

Unifying Frameworks for Nonlocality and Contextuality

Nadish de Silva*

*Department of Computer Science, University College London,
WC1E 6BT London, United Kingdom*

There are two frameworks, due to Cabello-Severini-Winter and Abramsky-Brandenburger-Hardy, for studying Bell and noncontextual inequalities; they differ vastly in their mathematical approaches. We unify their approaches and give simple, graph-theoretic methods for finding inequality-free proofs of nonlocality or contextuality and for finding states exhibiting strong nonlocality or contextuality. Finally, we apply these methods to concrete examples in stabilizer quantum mechanics relevant to understanding contextuality as a resource in quantum computation via magic state distillation.

I. INTRODUCTION

Bell famously showed that it is impossible to account for the empirical predictions of quantum mechanics with a model which is classical in the sense of being locally causal, i.e. one in which, upon conditioning on a common causal past, the joint probability distributions describing outcomes of experiments performed at different locations factorize into distributions associated with each location [1]. He derived the first *Bell inequality*, a bound on weighted sums of expectation values, that is satisfied by any locally causal model but violated by quantum theory. Entanglement and nonlocality remained insights of purely physical and philosophical significance for decades. Once studied by researchers trained in information processing, they became the foundation of novel, intrinsically quantum protocols for communication such as superdense coding, quantum teleportation, and quantum cryptography [2–4].

There are degrees of strength of nonlocality. The Bell state can be shown to violate a Bell inequality after repeated trials of an experiment. The Hardy state, however, is more nonlocal in the sense that it can demonstrate the impossibility of a locally causal model based on logical considerations without resort to inequalities. The GHZ state is even more nonlocal in that it gives the maximal possible violation of a Bell inequality.

Nonlocality is a special case of contextuality¹: the notion, established by Kochen and Specker [5], that the observable properties of a system cannot be assigned predetermined outcomes in a manner independent of the method of observation. Contextuality admits a description in terms of noncontextual inequalities, analogous to Bell inequalities, e.g. Klyachko et al.’s inequality [6].

Nonlocality has been studied abstractly as a resource for communication tasks [7] and it has been recognized that stronger nonlocality yields greater advantages, i.e. access to PR boxes allows for instantaneous communication [8]. Contextuality has recently been investigated as a resource in quantum computation. Howard et al. [9]

showed that contextuality is necessary in the magic state distillation (MSD) protocol for achieving universal quantum computation while Raussendorf [10] showed it is necessary for promoting the power of certain measurement-based quantum computers. Stronger contextuality correlates with greater computational gains. Clarifying the strengths of contextuality is needed to understand contextuality as a resource in quantum computation.

The results on contextuality as a resource rely on frameworks developed by Cabello-Severini-Winter (CSW) [11] and Abramsky-Brandenburger (AB) [12] which promote the example-based understanding of nonlocality and contextuality to a higher-level, structural one and expose the common structure of nonlocality and contextuality. They share the goal of providing tools for the analysis of nonlocality and contextuality in any experimental scenario, including a complete set of Bell or noncontextual inequalities. However, they differ vastly in their mathematical approach and how they relate to one another is unclear. We unify the two approaches and give simple, graph-theoretic methods for finding inequality-free proofs of nonlocality or contextuality and for finding states exhibiting strong nonlocality or contextuality. We then apply our methods to concrete examples coming from contextuality as a resource in MSD and find that certain magic states do not exhibit logical contextuality with respect to two-qutrit stabilizer operations.

II. FRAMEWORKS FOR NONLOCALITY & CONTEXTUALITY

In both the AB and CSW approaches, operational data is presented as likelihoods associated to outcomes of experiments which differ from classical probability theory in that it is not assumed that an experiment can simultaneously yield outcomes for all observable quantities; their mathematical presentations of operational data differ significantly. The dichotomy of locality versus nonlocality is the special case of the dichotomy of noncontextuality versus contextuality where the compatibility of observables is due to space-like separation of experimental sites.

Before presenting our unified approach, we outline the structures used in each framework and some key results.

* nadish.desilva@utoronto.ca

¹ We focus in this work on Kochen-Specker’s notion of noncontextual models: ones which satisfy outcome determinism.

A. The CSW graph-theoretic framework

The Cabello-Severini-Winter (CSW) framework employs the mathematics of graph theory² and combinatorial optimization to characterize nonlocality and contextuality for very general experimental scenarios. An experiment is formalized by an *exclusivity graph*. The vertices of such a graph describe events which give maximal information. Vertices representing mutually exclusive events are connected by an edge.

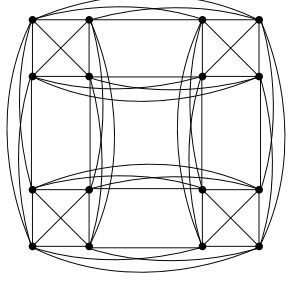


FIG. 1. The exclusivity graph for the standard Bell scenario. The four corner squares represent the events of an experiment with fixed choices of Alice's and Bob's settings. The edges between squares connect events where at least one party has chosen the same setting but observed a different outcome.

A preparation of a system yields probabilities p_i for each event. Every noncontextual inequality can be written as a bound on linear combinations of probabilities of events: $\sum_i w_i p_i$ where p_i is the likelihood of the i -th event and $w_i \geq 0$ is a coefficient. Such a linear combination is represented as a weighted graph (G, w) .

CSW showed that the independence number resp. Lovász theta number resp. fractional packing number are upper bounds for $\sum_i w_i p_i$ over all classical resp. quantum resp. locally orthogonal [13] models for the experiment represented by G :

$$\sum_i w_i p_i \leq \alpha(G, w) \leq \vartheta(G, w) \leq \alpha^*(G, w)$$

The bound $\sum_i w_i p_i \leq \alpha(G, w)$ is tight and every Bell or noncontextuality inequality is essentially of this form. A state is local resp. noncontextual precisely when it satisfies all such Bell resp. noncontextuality inequalities.

B. The AB sheaf-theoretic framework

In the AB framework³, an experiment is described not by an exclusivity graph but by a *measurement scenario*: a pair $(\mathcal{M}, \mathcal{C})$ where \mathcal{M} is a set of observables

and $\mathcal{C} \subset \mathcal{P}(\mathcal{M})$ is a family of subsets of \mathcal{M} such that $\cup C = \mathcal{M}$ and $C' \not\subset C$ for any $C, C' \in \mathcal{C}$. The set \mathcal{C} of contexts represents the maximal collections of compatible measurements or, equivalently, maximally refined experiments⁴. Measuring an observable yields a value from the outcome set \mathcal{O} usually given by $\{0, 1\}$.

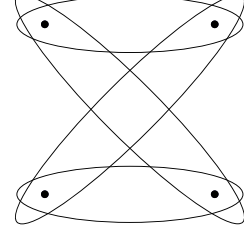


FIG. 2. The standard Bell scenario is described by two measurements for each of Alice and Bob, i.e. $X = \{A_1, A_2, B_1, B_2\}$ and contexts $C = \{\{A_i, B_j\} : i, j = 1, 2\}$ with $\mathcal{O} = \{0, 1\}$.

The *event sheaf* is the family of partial event sets $E(S) = \{e : S \rightarrow \mathcal{O}\}$ for each subset of measurements $S \subset \mathcal{M}$ together with course-graining maps $\gamma_{S'}^S(e) = e|_{S'}$ for each $S' \subset S$. The *distribution presheaf* is the family of sets $D(S)$ of probability distributions on events $E(S)$ together with maps $\mu_{S'}^S$, that take a distribution on $E(S)$ to its marginal distribution on $E(S')$ when $S' \subset S$.

A preparation of a system yields probability distributions for each experimental procedure, i.e. distributions ρ_C on $E(C)$ indexed by contexts $C \in \mathcal{C}$. Such data is *non signalling* when $\rho_C(o|M) = \rho_{C'}(o|M)$ for any pair C, C' of contexts, measurement $M \in C \cap C'$, and $o \in \mathcal{O}$; in this case, the distributions ρ_C constitute a *state* ρ .

A state is local resp. noncontextual precisely when its predictions can be accounted for by a locally causal resp. noncontextual hidden variable model. The hidden variable space can be taken to have the canonical form of $\Lambda = E(\mathcal{M})$ ⁵. Thus, a state ρ is local or noncontextual when there is a distribution $d \in D(\mathcal{M})$ which marginalizes to ρ_C for each context, i.e. $\rho_C = \mu_C^{\mathcal{M}}(d)$.

Abramsky-Brandenburger give logical descriptions of strengths of nonlocality and contextuality; Abramsky-Hardy give a complete set of Bell and noncontextual inequalities using logical consistency conditions. We detail both aspects of the framework below.

² Relevant basic definitions are found in Appendix A

³ We give an elementary presentation; it is originally phrased in terms of the mathematical language of *sheaf theory*.

⁴ Up to operational equivalence

⁵ That a state with any (outcome deterministic) noncontextual model can be accounted for by a canonical model is a vast generalization of Fine's theorem to all experimental scenarios.

III. A UNIFIED FRAMEWORK

An *observable event* is an $e \in E(C)$ for some context $C \in \mathcal{C}$. They are the elementary events observable in an experiment, i.e. the most refined propositions answerable by an experiment. Given a state ρ , we call an observable event e *impossible* when $\rho_C(e) = 0$; otherwise, it is *possible*.

Definition 1 The exclusivity graph $\mathcal{G}(\mathcal{M}, \mathcal{C})$ of the measurement scenario $(\mathcal{M}, \mathcal{C})$ has vertices given by the observable events $\{e : C \rightarrow \mathcal{O} \mid C \in \mathcal{C}\} = \cup_{C \in \mathcal{C}} E(C)$. Vertices $o_1 : C_1 \rightarrow \mathcal{O}$ and $o_2 : C_2 \rightarrow \mathcal{O}$ share an edge whenever $o_1|_{C_1 \cap C_2} \neq o_2|_{C_1 \cap C_2}$.⁶

Definition 2 The state graph of a state ρ is the graph \mathcal{G}_ρ which arises as the induced subgraph of $\mathcal{G}(\mathcal{M}, \mathcal{C})$ given by retaining only the possible events.

A. Hardy's paradox & logical contextuality

Considerable work has been done on the subject of establishing Bell's theorem without inequalities [15]. The idea is to preclude locally causal models of quantum theory using knowledge of the mere possibility or impossibility of certain correlated events rather than the precise correlations thereby yielding more robust experiments. While the first proof of Bell's theorem without inequalities was the GHZ state, Hardy's paradox [16] better exemplifies the subtle logical contradictions at play. As inequality-free proofs of nonlocality cannot be accomplished with standard Bell states, Hardy's paradox indicates a stronger form of nonlocality than the original one. In the AB framework, this notion is generalized to *logical nonlocality and contextuality*⁷; we exploit this to give a graph-theoretic characterization of operational data which preclude local or noncontextual models without inequalities.

A state is logically nonlocal resp. contextual if any locally causal resp. noncontextual model which can account for all the empirically observed events must also predict the occurrence of an event which is not empirically observed. This is described for a state ρ formally as follows: there is a possible event e such that any compatible hidden variable $\lambda \in \Lambda$ (one whose course-graining $\gamma_C^{\mathcal{M}}(\lambda)$ is e) must also predict the occurrence of an impossible event (there is a context $C' \in \mathcal{C}$ for which $\gamma_{C'}^{\mathcal{M}}(\lambda)$ is impossible).

Finding a graph-theoretic characterization of logical contextuality requires defining a graph invariant heretofore unused in the CSW approach:

Definition 3 A The independence degree of a vertex v in a finite graph G is the size of the largest independent set in G that contains v ; the minimal independence number is the minimum independence degree over all vertices.

Theorem 1 A state ρ is logically nonlocal or contextual (i.e. admits an inequality-free proof of nonlocality or contextuality) if and only if the minimal independence number of its state graph \mathcal{G}_ρ is less than the number of contexts.⁸

B. Maximal nonlocality & strong contextuality

The notion of *maximal nonlocality* was introduced by Elitzur-Popescu-Rohrlich [18] for states in a Bell experiment and generalized to all states by Barrett et al [19]. Abramsky-Brandenburger generalized this notion to maximal contextuality and gave an equivalent logical characterization called *strong contextuality* which we exploit to give a simple graph-theoretic criterion.

Some prominent examples of maximally nonlocal states include GHZ states (which are quantum states) and PR boxes (whose correlations are nonsignalling but cannot be realized within quantum mechanics). There are known advantages for communication tasks conferred by access to maximally nonlocal resource states. For example, sharing an unlimited number of PR boxes between two parties renders all communication complexity problems trivial.

Strong contextuality is known to play a role in quantum computation: Raussendorf demonstrated that strong contextuality is necessary for allowing certain measurement-based quantum computers to deterministically compute nonlinear functions.

A state ρ can be written as a convex decomposition into a noncontextual part A and a nonsignalling part Z :

$$\rho = \tau A + (1 - \tau)Z$$

by which we mean that, for all contexts $C \in \mathcal{C}$, the probability distributions ρ_C and $\tau A_C + (1 - \tau)Z_C$ in $D(C)$ are equal. The *noncontextual fraction* of ρ is the supremum of τ over all such convex decompositions. This concept generalizes the *local fraction* defined by Barrett et al. A state is *maximally contextual* when its noncontextual fraction is zero. This is equivalent to strong contextuality: there exists no hidden variable $\lambda \in \Lambda$ such that, for all contexts $C \in \mathcal{C}$, the observable event $\gamma_C^{\mathcal{M}}(\lambda)$ is possible.

Theorem 2 A state ρ is strongly nonlocal or contextual if and only if the independence number of its state graph \mathcal{G}_ρ is less than the number of contexts.⁸

⁶ c.f. the hypergraph construction of measurement protocols [14]

⁷ Also known as *possibilistic* nonlocality or contextuality

⁸ The proof is found in Appendix B

C. Logical Bell inequalities

The *logical Bell inequalities* of $(\mathcal{M}, \mathcal{C})$ are a distinguished class of Bell or noncontextual inequalities identified by Abramsky and Hardy [20] which are derived from logical consistency conditions. Given N Boolean formulas f_i with variables drawn from \mathcal{M} that are assigned likelihoods p_i of being true, $\sum p_i \leq N - 1$ whenever the formulas are logically inconsistent. A logical Bell inequality is built by choosing a subset of the observable events $E_i \subset E(C_i)$ from each context $C_i \in \mathcal{C}$ and constructing the formulae:

$$f_i = \bigvee_{e \in E_i} \bigwedge_{M \in C_i} \begin{cases} M & \text{if } e(M) = 1 \\ \neg M & \text{if } e(M) = 0 \end{cases}.$$

When these formulas contain a contradiction, we obtain an inequality: the sum of the likelihoods of the events in $\cup E_i$ is less than or equal to $N - 1$.

A state is logically nonlocal or contextual precisely when it violates a logical Bell inequality. A state is strongly nonlocal or contextual precisely when it maximally violates a logical Bell inequality: $\sum p_i = N$.

Logical Bell inequalities can be slightly extended by allowing integer coefficients: $\sum n_i p_i \leq K$. Abramsky and Hardy give an algorithm for generating a complete description of the local or noncontextual polytope using inequalities equivalent to these.

Both standard and extended logical Bell inequalities can be derived as CSW inequalities on the exclusivity graph $\mathcal{G}(\mathcal{M}, \mathcal{C})$. Standard inequalities are constructed by choosing the sets of observable events E_i as above and giving them weight one while all other events are given weight zero. Extended inequalities are constructed by giving the events of E_i the non-negative integer weight n_i and all other events weight zero. In both cases, the upper bound is given by the independence number.

IV. APPLICATIONS

In this section, we outline some example computations which yield insight into the contextuality of quantum states relative to stabilizer operations. Recent work has shown contextuality with respect to two-qudit stabilizer operations to be a necessary condition for MSD in odd-prime-power qudit systems.

A. Qubit stabilizer operations

It is impossible to use mere contextuality as a criterion for determining which qubit states are suitable for MSD. This is because the argument of Peres-Mermin [21] identifies all two-qubit states as contextual with respect to two-qubit stabilizer operations. One might hope to use stronger forms of contextuality as a criterion for identifying possible resource states for two-qubit MSD. However,

it turns out that a straightforward extension of the Peres-Mermin argument identifies all two-qubit states, including the maximally mixed state, as not only contextual but strongly contextual.

Result 1 *The n -qubit maximally mixed state is strongly contextual with respect to n -qubit stabilizer operations whenever $n > 1$.⁸*

This is an example of the general phenomenon of *state-independent (strong) contextuality* which occurs whenever the independence number of the orthogonality graph of a quantum measurement scenario is less than the number of contexts (AB give a large family of examples of measurement scenarios exhibiting state-independent strong contextuality in any quantum realization [12, §7]; c.f. [22]).

B. Qudit stabilizer operations

In the qudit cases where contextuality is a useful necessary criterion for identifying MSD resource states, it is not known whether it is a sufficient criterion. This opens the question of whether strengths or measures of contextuality can be correlated with distillability. We find, surprisingly, that although qutrit magic states can demonstrate maximal violation of some CSW noncontextual inequalities, it exhibits a low degree of contextuality relative to general nonsignalling states.

Result 2 *The states $|M\rangle \otimes |M\rangle$, where $|M\rangle$ is the qutrit magic state $|M_0\rangle$ [23], $|E\rangle$, $|N'\rangle$ [24], or $|\psi_{U_n}\rangle$ [25], are neither logically nor strongly contextual with respect to two-qutrit stabilizer operations.*

Following Howard et al., we consider two-qutrit stabilizer operations. The state graph of any two-qutrit quantum state is an induced subgraph of the two-qutrit orthogonality graph. The latter was computed using Gross' [26] phase space formulation by enumerating the 40 Lagrangian subspaces of $(\mathbb{Z}_3 \times \mathbb{Z}_3)^2$. Finally, the corresponding state graphs and their minimal independence numbers were computed and found to be 40.

V. CONCLUSIONS

We have shown how to unify the two leading quantitative frameworks, due to Cabello-Severini-Winter and Abramsky-Brandenburger, for nonlocality and contextuality of operational data for very general experimental settings by constructing an exclusivity graph for any measurement scenario. In doing so, we clarify CSW's notion of *event* and *mutual exclusivity* which were heretofore given an explicit meaning only once interpreted within a particular theory.

We then exploit this unification to give simple, graph-theoretic characterizations of states which admit

inequality-free proofs of nonlocality or contextuality (i.e. logical nonlocality or contextuality) and of states exhibiting maximal nonlocality (in the sense of Elitzur-Popescu-Rohrlich) or strong contextuality. As a further consequence, the logical Bell inequalities of Abramsky-Hardy are derived as examples of CSW inequalities.

Finally, we apply our methods to contextuality with respect to stabilizer operations. Our result first precludes the use of strengths of contextuality as a means of analyzing qubit MSD protocols which make use of all stabilizer operations. Our second result shows that while strong contextuality is a necessary resource for certain measurement-based models of quantum computation, it is not necessary for qutrit MSD.

Apart from the obvious applications towards generating robust experimental tests of nonclassicality, a tractable classification of strengths of contextuality is

necessary for understanding the role of quantum resources in information processing tasks.

Further development will require developing natural quantitative measures of contextuality, investigating the relationship between strengths of contextuality and other models of quantum computation, and understanding the relationship with Spekkens' [27] notion of contextuality which relaxes the assumption of outcome determinism.

ACKNOWLEDGMENTS

The author wishes to thank Eric Cavalcanti, Samson Abramsky, Simone Severini, Rui Soares Barbosa, Kohei Kishida, Dan Browne, Shane Mansfield, and Andrew Simmons for many helpful discussions. This research was financially supported by the EPSRC.

-
- [1] J. S. Bell, *Physics* **1**, 195 (1964)
 - [2] C. Bennett and S. Wiesner, *Phys. Rev. Lett.* **69**, 2881 (1992)
 - [3] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *Phys. Rev. Lett.* **70**, 1895 (1993)
 - [4] C. H. Bennett and G. Brassard, in *International Conference on Computer System and Signal Processing, IEEE, 1984* (1984) pp. 175–179
 - [5] S. Kochen and E. P. Specker, *J. Math. Mech.* **17**, 59 (1967)
 - [6] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, *Phys. Rev. Lett.* **101**, 020403 (2008)
 - [7] J. Barrett, *Phys. Rev. A* **75**, 032304 (2007)
 - [8] W. van Dam, *Nonlocality & communication complexity*, Ph.D. thesis, Faculty of Physical Sciences, University of Oxford (1999)
 - [9] M. Howard, J. Wallman, V. Veitch, and J. Emerson, *Nature* **510**, 351 (2014)
 - [10] R. Raussendorf, *Phys. Rev. A* **88**, 022322 (2013)
 - [11] A. Cabello, S. Severini, and A. Winter, *Phys. Rev. Lett.* **112**, 040401 (2014)
 - [12] S. Abramsky and A. Brandenburger, *New J. Phys.* **13**, 113036 (2011)
 - [13] T. Fritz, A. B. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acín, *Nature Comm.* **4** (2013)
 - [14] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, *Comm. Math. Phys.* **334**, 533 (2015)
 - [15] D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger, *Am. J. Phys* **58**, 1131 (1990)
 - [16] L. Hardy, *Phys. Rev. Lett.* **68**, 2981 (1992)
 - [17] R. B. Allan and R. Laskar, *Disc. Math.* **23**, 73 (1978)
 - [18] A. C. Elitzur, S. Popescu, and D. Rohrlich, *Phys. Lett. A* **162**, 25 (1992)
 - [19] J. Barrett, A. Kent, and S. Pironio, *Phys. Rev. Lett.* **97**, 170409 (2006)
 - [20] S. Abramsky and L. Hardy, *Phys. Rev. A* **85**, 062114 (2012)
 - [21] N. D. Mermin, *Rev. Mod. Phys.* **65**, 803 (1993)
 - [22] A. Cabello, M. Kleinmann, and C. Budroni, *Phys. Rev. Lett.* **114**, 250402 (2015)
 - [23] H. Anwar, *Towards fault-tolerant quantum computation with higher-dimensional systems*, Ph.D. thesis, University College London (2014)
 - [24] H. Dawkins and M. Howard, *Phys. Rev. Lett.* **115**, 030501 (2015)
 - [25] M. Howard and J. Vala, *Phys. Rev. A* **86**, 022316 (2012)
 - [26] D. Gross, *J. Math. Phys.* **47**, 122107 (2006)
 - [27] R. W. Spekkens, *Phys. Rev. A* **71**, 052108 (2005)

A. GRAPH THEORY

A (simple, undirected) *graph* G is a set of vertices $V = \{v_1, \dots, v_n\}$, pairs of which may be joined by an edge. The edges are represented by a set $E \subset V \times V$ such that $(v, v) \notin E$ for any vertex v and such that $(v, v') \in E$ if and only if $(v', v) \in E$. That is, no edge connects a vertex to itself and there is an edge between v and v' whenever there is an edge between v' and v . Given a subset $V' \subset V$ of vertices, the *induced subgraph* has V' as vertices and retains all edges $(v, v') \in E$ with $v, v' \in V'$.

A weighted graph (G, w) is a graph G together with a function $w : V \rightarrow \mathbb{R}^{\geq 0}$ from vertices to non-negative reals. We denote the weight $w(v_i)$ of a vertex v_i by w_i .

An *independent (vertex) set* is a subset $I \subset V$ such that no two vertices in I share an edge: $(v, v') \notin E$ whenever $v, v' \in I$. A *clique* is a subset $C \subset V$ such that every two distinct vertices in C share an edge: $(v, v') \in E$ whenever $v, v' \in C$ and $v \neq v'$.

The *independence number* of a graph G is the size of the largest independent set in G . The independence number of a weighted graph (G, w) is the maximum value of the sum $\sum_{i \in I} w_i$ where I is any independent set.

To define the *Lovász theta number* $\vartheta(G, w)$ of a weighted graph (G, w) we must first define an orthonormal representation of a graph G . This is a choice of a unit vector $|\phi\rangle \in \mathbb{R}^d$ and an assignment to each vertex

v_i a unit vector $|\psi_i\rangle \in \mathbb{R}^d$ such that $\langle \psi_i | \psi_j \rangle = 0$ whenever v_i and v_j share an edge. The choice of dimension d can be arbitrary. The Lovász theta number $\vartheta(G, w)$ is then defined to be the maximum $\sum_{i \in V} w_i |\langle \psi_i | \phi \rangle|^2$ over all possible orthonormal representations of G .

The *fractional packing number* $\alpha^*(G, w)$ is the maximum possible value of the sum $\sum_{i \in V} p_i w_i$ where $\{p_1, \dots, p_n\}$ is a choice of a non-negative real number for each vertex such that $\sum_{i \in C} p_i \leq 1$ for every clique $C \subset V$.

B. PROOFS

Theorem 1 *A state ρ is logically nonlocal or contextual (i.e. admits an inequality-free proof of nonlocality or contextuality) if and only if the minimal independence number of its state graph \mathcal{G}_ρ is less than the number of contexts.*

Suppose a state ρ gives rise to a state graph \mathcal{G}_ρ whose minimal independence number is less than $|\mathcal{C}|$. This means that there is a possible event e with independence degree less than $|\mathcal{C}|$. If $\lambda \in \Lambda$ is a compatible hidden variable, then the observable events $\{\gamma_{C'}^M(\lambda) | C' \in \mathcal{C}\}$ clearly form an independent subset of $\mathcal{G}(\mathcal{M}, \mathcal{C})$ of size $|\mathcal{C}|$. One of these events cannot be in the state graph \mathcal{G}_ρ and, so, is impossible.

Conversely, suppose that a state ρ gives rise to a state graph \mathcal{G}_ρ whose minimal independence number is $|\mathcal{C}|$ and e is a possible event. (The minimal independence number cannot be larger as the sets $E(C)$ form a clique cover of $\mathcal{G}(\mathcal{M}, \mathcal{C})$.) The event e must be contained in an independent subset I of \mathcal{G}_ρ of size $|\mathcal{C}|$. Construct the hidden variable $\lambda : \mathcal{M} \rightarrow \mathcal{O}$ by $\lambda(M) = i(M)$ for some $i \in I$ whose context contains M . This is well-defined by virtue of the fact that I is an independent set. As the sets $E(C)$ form a clique cover of $\mathcal{G}(\mathcal{M}, \mathcal{C})$, the domains of i cover \mathcal{M} . Clearly, each i is the course-graining of λ and is possible.

Theorem 2 *A state ρ is strongly nonlocal or contextual if and only if the independence number of its state graph \mathcal{G}_ρ is less than the number of contexts.*

If the state graph \mathcal{G}_ρ of a state ρ contains an independent set I of size $|\mathcal{C}|$, then, as in the previous section, construct the hidden variable $\lambda : \mathcal{M} \rightarrow \mathcal{O}$ by $\lambda(M) = i(M)$ for some $i \in I$ whose context contains M . Clearly, each i is the course-graining of λ and is possible.

Conversely, if there exists $\lambda \in \Lambda$ such that $\gamma_C^M(\lambda)$ is possible for all contexts $C \in \mathcal{C}$, then the $\gamma_C^M(\lambda)$ form an independent subset of \mathcal{G}_ρ of size $|\mathcal{C}|$.

Result 1 *The n -qubit maximally mixed state is strongly contextual with respect to n -qubit stabilizer operations, whenever $n > 1$.*

Consider the measurement scenario consisting of those n -qubit Pauli operators whose square is the identity as measurements and, as contexts, those subsets which generate maximal abelian subgroups of the n -qubit Pauli group P_n . The state graph of the maximally mixed state, i.e. those formal events which are possible within quantum theory, is the orthogonality graph of n -qubit stabilizer quantum mechanics. It has rank-1 stabilizer projections as vertices with edges joining orthogonal projections. We must show that the independence number of this graph is less than the number of contexts $|C_n|$.

Suppose for contradiction that I is an independent vertex subset of size $|C_n|$. I must contain precisely one state from each context for otherwise it will contain two distinct eigenstates for the Paulis of the context which must be orthogonal and thus connected by an edge. We denote the stabilizer state chosen from a context C by I_C .

Thus, I yields a noncontextual value assignment $v : P_n \rightarrow \{\pm 1, \pm i\}$. For an n -Pauli $P \in P_n$ in a context $C \in C_n$, $v(P)$ is eigenvalue associated by P to the state I_C . This eigenvalue is independent of the choice of context for if P is a member of both C and C' , then it associates the same eigenvalue to both I_C and $I_{C'}$; otherwise, I_C and $I_{C'}$ belong to different eigenspaces of P which contradicts the assumption that no pair from I is orthogonal. We can extend v to all Paulis by multiplicativity. However, there exists no function $v : P_n \rightarrow \{\pm 1, \pm i\}$ mapping Paulis to one of their eigenvalues such that $v(-P) = -v(P)$ and $v(PP') = v(P)v(P')$ whenever P and P' commute.

We first reproduce the standard Mermin argument for 2-qubit systems. Consider the table of Paulis:

$Z \otimes I$	$I \otimes X$	$Z \otimes X$
$I \otimes Z$	$X \otimes I$	$X \otimes Z$
$Z \otimes Z$	$X \otimes X$	$Y \otimes Y$

Each entry has eigenvalues in the set $\{\pm 1\}$. The rows and columns give commuting triples of P_2 . Suppose a valuation v satisfying our hypotheses exists and consider the product of $v(P)$ over all P in each row or column. Since each entry appears in one row and one column, this is the product of $(\pm 1)^2$ nine times: 1. However, if we collect the terms into rows and columns, we find that we are taking the product of $v(I \otimes I)$ five times with $v(-I \otimes I)$ once. Therefore, the product is -1 and we reach a contradiction.

For the general $n > 2$ case, we repeat the same argument only we tensor all of the entries of the table with $I^{\otimes(n-2)}$ to get entries in P_n . This does not change the eigenvalues of the entries as I has only eigenvalue 1 nor does it affect the commutativity of the rows and columns. The product of the rows and columns yields $I^{\otimes n}$ five times and $-I^{\otimes n}$ once.